Appendix B

Product Requirements

IPSX IP Flow Cache Architecture

3277786

Number: CCC-NNNNN-RR

Version: 1.0

Date: August 6, 2000

Author's Name(s):

Abraham Rabindranath Mathews, Kevin Lin,
Kiho Yum, Naveed Alam, Podibanda Kurrupu

Reviewers

	Name	Check
1	Manager	√
2	Project Lead	√
3	Reviewer 1	√
4	Reviewer 2	√

CoSine COMMUNICATIONS

Document History

Version	Date	Author	Change Description
1.0	8/6/2000	Your Name	Establish document.

Table of Contents

1 Introduction

This document is intended to describe main features of IPSX IP Flow Cache Architecture. The cache system is designed to reduce the processing overhead associated with the IP traffic passing through the system. The flow cache based forwarding helps in achieving better and more deterministic throughput and latency goals. It also helps the system resources like CPU, Memory and Interconnect Bandwidth to scale better as traffic load increases through the system.

There are three major areas identified in IFCA that have contributed the most in achieving the required performance and scalability goals.

1- Multi Service Forwarding Cache

2- Multiple Cache Lookups per Flow

3- Distributed Flow Cache

2 Background

There is another document by Abraham Mathews "IPSX Packet Processing Requirements" that helps in understanding the underlying problem and suggested solutions.

2.1 Terms and Acronyms

IPSec	Protocol for providing security at the IP Layer
L2TP	Layer 2 Tunneling Protocol that runs on top of UDP layer
GRE	Generic Encapsulation, a layer 4 protocol
NAT	Network Address Translation
SPF	Stateful Packet Filter
PE	Processing Elements in IPSX System
IPNOS	IP Network Operating System
VR	Virtual Router
IFCA	IP Flow Cache Architecture
IPSX	Cosine Communications IP Service Switch 9000

2.2 Glossary

L3 Flow	Only Layer 3 attributes are used to define the flow. The attributes used are <Destination IP Address, Source IP Address, Type of Service, IP Protocol>
L4 Flow	Attributes are used from layers beyond L3 to define the flow. Some protocols that contribute in L4 flows are TCP, UDP, ICMP, GRE and IPSec. In case of TCP or UDP the Layer 4 attributes used are <Destination Port, Source Port>
Packet Flow	General term used for L3 or L4 flows depending on the attribute types used.
Strict L3 Flow	flow consists of IP fragments; L4 information is not present in all fragments
IP Forwarding Services	Represent processing requirements for the ingress and egress packet flows in IPSX. The basic services consist of Routing and Forwarding services. The extended services consist

	of VPN and security related services like NAT, L2TP, Firewall, IPSec, SPF etc. These services are optional and can be selectively enabled by the IPSX user.
Host	IPSX itself, Host addressed packets are incoming packets with destination IP address as one of the IPSX's owned IP address
Configured Topology	Given a Subscriber VR and its ISP VR, consider the associated graph. This graph's nodes are the PE's executing software/ services and the edges of the graph are the communication links between the software/ services. This graph is the configured topology.
Flow Topology	Based on the actual dynamic flow of L3 and L4 packets a graph of software/ services and the communication links between the software/ services can be constructed. This graph is the flow topology. The flow topology is derived from and would be a sub-graph of the configured topology if shortcuts are not employed. When shortcuts are employed, there is no direct graph theoretic correlation between the configured and flow topologies.

3 Multi-Service Cache Design

The IPSX Packet Forwarding goes through complex set of services. The basic or mandatory services consist of Routing and Forwarding services. Traditional Routers normally cache the necessary information about basic services that help in forwarding L3 flows with reduced processing overhead. If the extended services are enabled, they bring their own processing overhead in the IP forwarding path as shown in Figure 1. The extended services are normally independently designed and they bring a wide variety of processing requirements. The ability to recognize packet flows and apply flow specific processing reduces lot of processing overhead associated with these services.

The IFCA allows extended services to create their own cache contexts during the initial learning phase of the packet flows. The cache context consists of all the necessary information needed to process a service. Once the learning phase is done, the cache entry gets filled with an ordered set of operations based on the pushed cache contexts and is inserted into the hash table. The following packets belonging to the same flow find the cache entry in the hash table and go through only cache-enabled operations. The IFCA is scalable and extensible to include cacheable support for new services as they get implemented in IPSX platform.

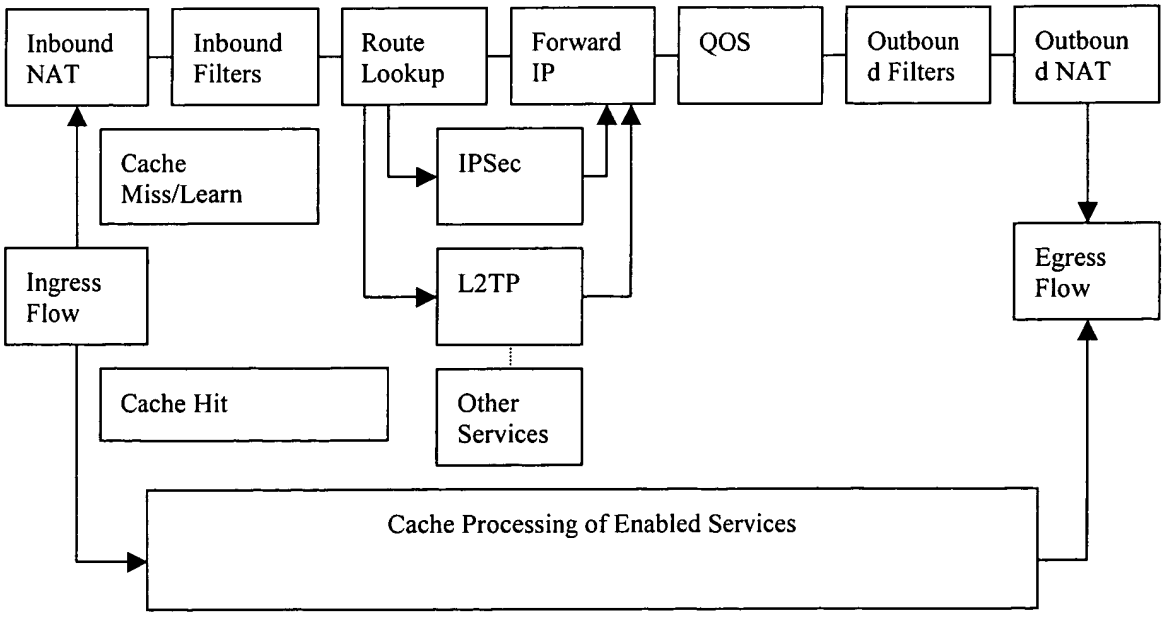

Figure 1

4 Multiple Cache Lookups for Complex Flows

The Strict L3 flows, tunneled flows and encrypted flows bring another level of complexity to the flow processing. During processing, the packet's L3 and L4 attributes transform from one type to the other. In order to keep the processing within the scope of cache operations, IFCA applies packet classifications multiple times as the L3 and L4 attributes change. This results in multiple cache lookups for the same ingress flow and enables VR to finish processing of packets through cache operations only. The traditional routers normally switch to slow path to process more complex packet flows.

As an example consider Ingress flow consisting of host addressed, encrypted, tunneled and fragmented packets. Figure 2 shows processing of Strict L3 ingress Flow of P1, P2 and P3 packets. The ingress flow goes through 3 cache lookups before turning into egress flow.

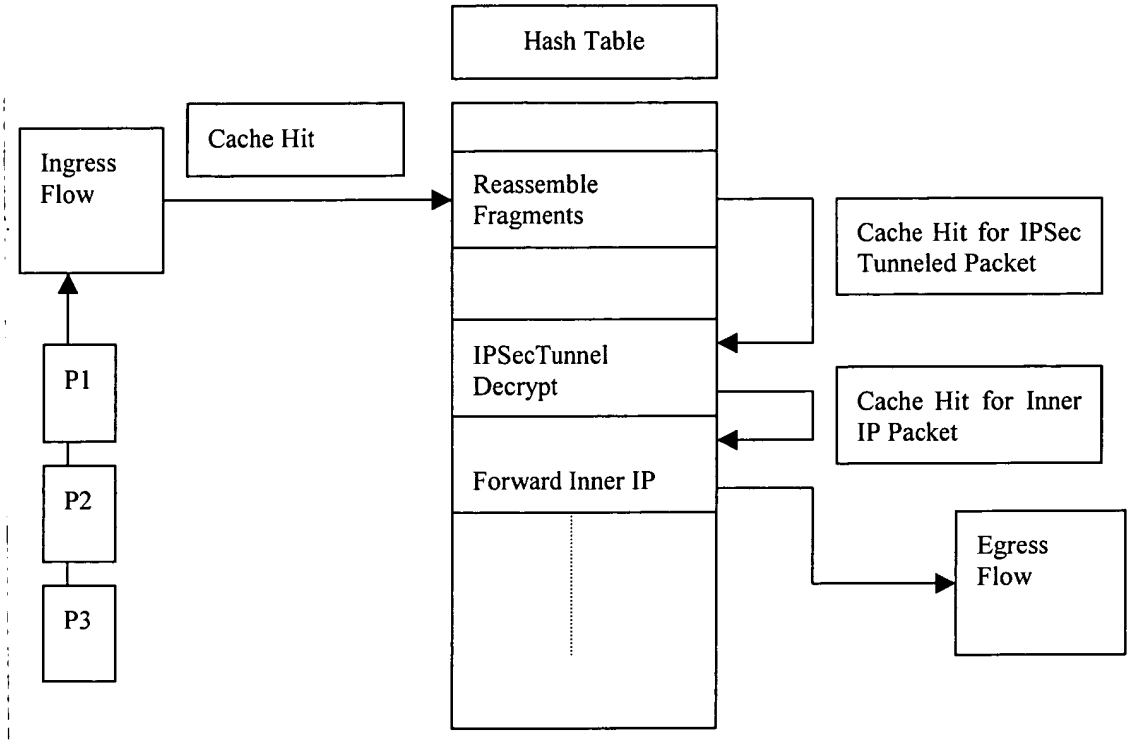

Figure 2

5 Distributed Flow Caches

IPSX provides cluster based distributed Processor Architecture. The multistage interconnect provides logical meshed topology for all PE-PE communication channels. Some PEs are equipped with special hardware like Crypto PEs to provide dedicated services. As the VR enables extended services, the IPNOS dynamically enables and links the services to the VR. Some of these services could end up on different PEs spread throughout the system. The VR cannot afford to limit its internal routing intelligence to one PE to process multiple services on different PEs. This can result into Star like flow topology putting more load on the resources of one PE doing all the internal routing for the VR.

In the process of applying services to the packet flows, the VR acquires sufficient knowledge of the internal routing of the flows. The IFCA allows the VR to take advantage of this knowledge and underlying meshed topology of PEs. The VRs can dynamically distribute the information about their flow topology to other VRs in the system. This results in optimum flow paths within the system where packet flows can directly reach the service PEs from any VR.

As an example consider two VRs, Access and ISP VR. The ISP VR is doing simple forwarding for Access VR. The Access VR has two services enabled, Intrusion Detection and IPSec. The configured topology in Figure 3 shows that Access VR has the initial knowledge of the enabled services.

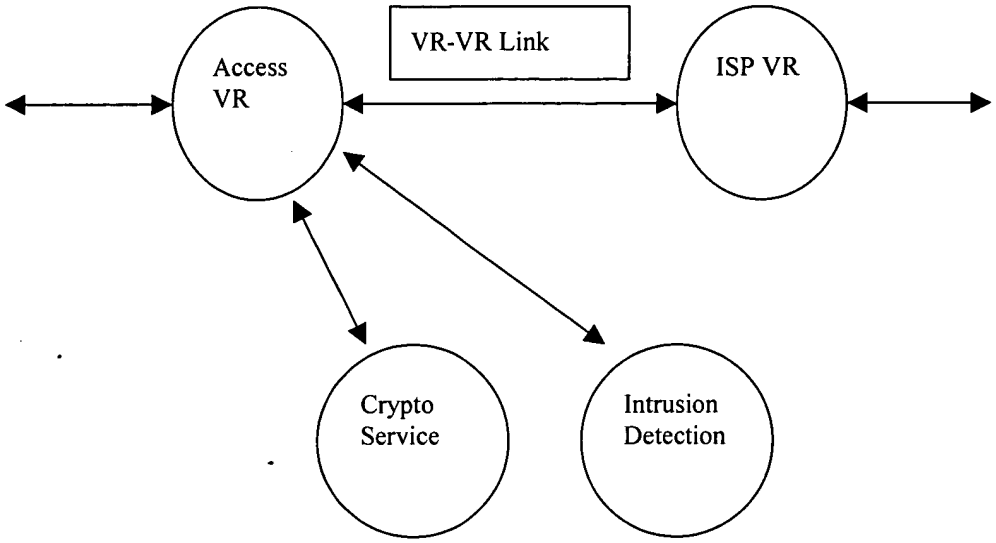

Figure 3

There are two L4 Flows coming to Access VR from ISP VR. The flow F1 needs Intrusion Detection only and F2 needs Intrusion detection as well as IPSec decryption. The Access VR sends flow redirection messages to ISP VR which inserts new entries for routing of F1 and F2 flow. This redefines flow topology for both the flows as shown in figure 4

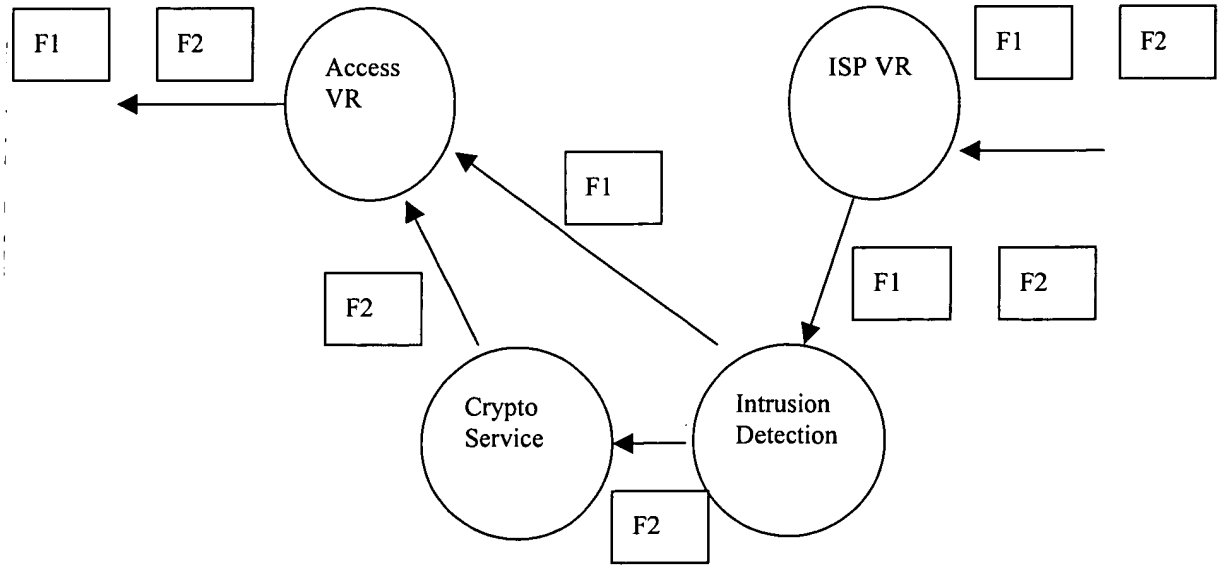

Figure 4